

## Arrays Cast Toxicology in a New Light

Out on the San Diego beaches in January 2000, the breezes were warm and soft, but inside the Hilton Resort, at a special session of the American Association for Cancer Research conference on DNA repair defects, a storm was boiling.

Leona Samson, director of the Biochemical and Experimental Toxicology Research Core at Harvard's Kresge Center for Environmental Health and a Harvard School of Public Health professor of toxicology, had just presented new data showing that the typical response to certain DNA-damaging agents by yeast cell gene expression was much more complicated than ever previously seen, or even suspected, and that many more genes were involved than previously imagined. "It made quite a splash," she recalls. "I was surrounded by a sea of people, and I remember one person saying, 'Oh my God, I should probably just give up what I'm doing.'"

Working at the Kresge Center, one of the NIEHS's 21 environmental health sciences centers, and using microarray technology, which allows scientists to watch an organism-wide response to DNA damage, Samson and her labmates found that exposure to alkylating agents, widely used as a model system for studying cellular responses to DNA damage, evoked at least threefold changes in expression in one-third of the yeast genome, or about 2,000 genes. The news came as a shock to scientists, who had believed that the response to alkylating agents was limited solely to DNA repair genes.

Even Samson herself, who has spent her entire career studying specific DNA repair mechanisms, was struck by her own results. "We were, in a sense, frightened at how big the response was," she recalls. "We said, 'What are we going to do with so many genes?'"

Genomics, the study of how cells' entire genetic complement responds to

environmental and internal stimuli, is rocking the boat in the field of toxicology. The impact of new waves of information has resonated throughout the field, and has more researchers talking about toxicogenomics. This approach to studying toxicology focuses on cellular changes in gene expression and could change the way toxicants are detected and classified.

Many researchers who would like to obtain the hardware themselves have been left out by the high cost of the technology. While genomic arrays are spreading relatively quickly through the industrial community, academic toxicologists have been less able to shoulder the considerable costs of using microarrays. A single 6,200-gene chip currently costs about \$1,000.

However, microarrays make it possible to discover in a matter of days or weeks what would have once taken years to do. Cells are grown, then split into control and treated groups. Isolated pieces of RNA are labeled with a fluorescent dye and then allowed to hybridize to chips, each of which is studded with up to 40 oligonucleotides complementary to each of the 6,200 specific yeast genes and open reading frames (areas that are likely to be protein-coding regions of the genome). A scanning confocal microscope can measure changes in fluorescence as low as twofold.

### Heading in a New Direction

In 1997, Samson and then-postdoctoral student Scott Jelinsky, now a staff scientist at Genetics Institute in Boston, Massachusetts, were discussing which line of investigation he should pursue when they heard about a user program offered by Affymetrix, a Santa Clara, California, company that makes the inches-square chips on which genomic studies are performed. Another Harvard laboratory, headed by George Church, already had much of the equipment needed to use the chips, so Jelinsky took two trips to California, where he learned to set up experiments and collect expression data.

Not long afterward, the Samson lab began its first genomic studies of the effects of methyl methanesulfonate on yeast, later published in the February 1999

issue of *Proceedings of the National Academy of Sciences*. This preliminary snapshot showed that expression of 325 genes was up-regulated after exposure to the alkylating agent, while 76 genes were down-regulated. This meant that far more genes were somehow involved in a response to DNA damage than scientists had previously believed.

Samson had spent most of her career focusing her attention on the response to alkylation of DNA repair genes such as *MAG1*. *MAG1*'s protein product initiates the removal and replacement of damaged DNA base pairs by recognizing and excising damaged bases. *MAG1* is found in a wide variety of species, including various plants, insects, fish, rodents, and humans.

That *MAG1* seemed lost in such a cacophony of gene expression suggested a much broader, more complicated response to cellular damage than Samson had ever imagined. "But when we sat down to write the paper," she recalls, "it dawned on me that we're flooding cells with a highly reactive chemical; of course it's going to change lots of things. These organisms are exquisitely tuned to sense what's going on in their environment. If you have a chemical that reacts with virtually every class of molecule in the cell, it's going to respond when things are changed. So I'm no longer surprised by the breadth of response."

Still, the task of interpreting the results seemed daunting. Hoping to shed some light on the rhyme and reason of all this activity, Samson and her colleagues decided to broaden the study to look at cells at various time points after exposure to a variety of alkylating agents, including methyl methanesulfonate. All in all, they performed 26 transcriptional profiles, which, as Samson notes, would have taken years with the standard Northern blot techniques they had been using only months earlier to do similar studies. When 2,000 yeast genes appeared to have been affected by the agents, Samson knew that she had to find some way to fathom the results. Her answer was to look back at *MAG1*.

"We tried to get our arms around it by asking a question from our own expertise," Samson says. "[What other gene] behaves the same as *MAG1*?' It was an obvious question."

That question turned out to have provocative answers. Thirty-three protein degradation genes, or proteasomes, along with four additional DNA repair genes

turned out to be coregulated with *MAG1*, meaning that the protein expression of the genes rises and fall together because they are regulated by the same genetic elements, or factors. It had never before occurred to Samson that protein degradation would walk hand-in-hand with DNA repair, but it was totally logical, she says, that faulty proteins would have to be removed before fresh ones could be expressed and installed.

"We had been doing very traditional reductionist biology and concentrating on repair pathways," she says. "But by looking globally with genome arrays, we were forced to open our eyes to what was there and ask questions about other genes."

The coregulation of DNA repair and protein degradation genes became even clearer when Jelinsky found that a genetic element regulating *MAG1* that they had identified eight years ago is identical to a regulator of the proteasomes. By reviewing and analyzing the overall expression response to toxicants, he and Samson had unearthed new dimensions of the damage response and found the responsible DNA machinery.

### A More Creative Approach

Formerly unimpressed with the vast amounts of data churned forth by genomic studies, researchers are now thinking more creatively about the analytic approaches. "We need to think about patterns and shapes in the data," says Samson. "I think in the end it will come back to individual genes. But first we can look at how a stimulus affects an organism's entire biology, and then we can begin dissecting the response."

Genomic arrays allow toxicologists to look at cellular behavior in a completely new light. In a sense, recording individual gene responses to powerful insults such as alkylating agents was akin to studying the effects of poverty by monitoring a person's bank account—the complete picture is much larger than what is actually being measured. But genomic arrays simultaneously report indicators of multiple dimensions of the cellular response to stimuli. Now, in addition to gaining insight into basic cellular mechanisms of repair, researchers looking at a variety of indicators and responses of toxicity may gain some predictive power regarding individual compounds—and individual humans.

Both academic and private laboratories have already begun work on finding genes that induce protection or sensitivity to tox-

icants in individual cells and people. Spencer Farr, founder and CEO of Phase-1 Molecular Toxicology, based in Santa Fe, New Mexico, says that genomic analysis of individual response to substances will challenge scientific assumptions about the nature of toxicity.

"Our experiments are based on the idea that there's an 'average person' that typifies our response to toxicants," says Farr. "The reality is that, no matter what toxicant or compound you're looking at, there are always going to be subpopulations that react differently from the norm."

Farr believes that toxicogenomics could revolutionize the study and implementation of drugs that cause dangerous side effects in defined patient subgroups. The psychoactive drug clozapine, for example, causes dangerous blood-clotting disorders in about 5% of patients. Toxicogenomic studies of vulnerable patients could lead to the identification of vulnerability genes or gene patterns that could be cheaply detected with mini-arrays. Farr's company is concentrating on gene expression in the liver, kidneys, and other organs that are often targeted by drugs.

Farr also believes toxicogenomics will become increasingly important in preclinical drug studies. "Rats predict carcinogenicity in mice with only about 50% accuracy," he points out. "How predictive do you think either of them are in human reactions to drugs? Using gene expression and analysis, I believe that we can do a better job of predicting who will suffer adverse outcomes from all kinds of substances."

A number of laboratories are also gearing up to begin studying how to predict the toxicity of a given substance by comparing patterns of evoked gene expression among new and known compounds. This could be particularly useful in gauging the potential impact of long-term exposures, or the effects of certain compounds on fetal development.

### Investing in the Future

Samson's lab is now looking for ways to lower the cost of doing toxicogenomics by making their own chips. Led by Thomas Begley, a postdoctoral fellow in her lab, Samson's staff is currently amplifying each of yeast's 6,200 genes and cross-linking them to lysine-coated slides. For an initial investment of about \$100,000, Begley says, the lab will make chips for several dollars apiece. Soon, they will be using them to

study gene expression in *Escherichia coli* and other model systems, including higher organisms such as knockout mice.

At this point, the greatest obstacle to the spread of toxicogenomics may remain in interpreting the multitudes of data that pour out of each slide. At this point, it's difficult to say what the crests and valleys of gene expression mean—whether an increase in the expression of a particular gene indicates involvement in protection or vulnerability to a toxicant.

"There's a waiting aspect to this," says Harvard professor of toxicology Armen Tashjian, Samson's predecessor as core director at the Kresge Center. "People are still looking for validation, and that's going to come with time and more experiments."

For her part, Samson feels that the time has come for more scientists to look at genomics as a guide to their research, rather than as a threatening alternative to current approaches of studying single or small numbers of genes. "These kinds of profiles have to be made available for everyone to look at," she says. "The data should be there for everyone. We're looking at the data from one simple point of view, but I think they need to be examined from each person's point of view, which relates to their own area of expertise." —John F. Lauer



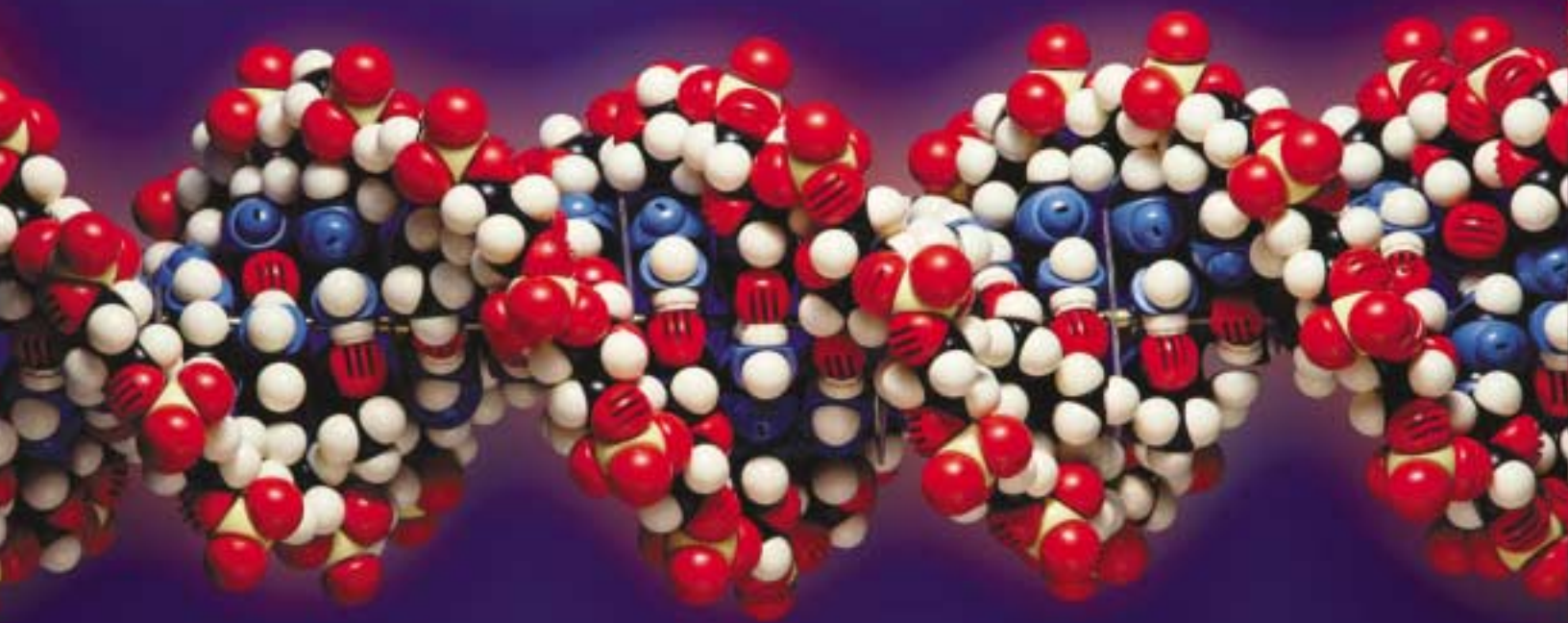
## New Center a Stroke of Gene-ius

NIEHS scientists are working with the Human Genome Project to make environmental medicine uniquely individual. On 7 December 2000, the NIEHS officially opened the National Center for Toxicogenomics (NCT) to begin studying how thousands of genes interact and respond to environmental exposures during different stages of health and disease.

Some of the diseases associated with environmental factors that toxicogenomics may help to elucidate include cancer, pulmonary disease, neurodegenerative disorders, developmental disorders, birth defects, reproductive dysfunction, and autoimmune disease. Toxicogenomics is a relatively new field that uses microarray technology and incorporates information from the Human Genome Project to develop highly individual toxicologic assays. Thus, says Ben Van Houten, the coordinator for the new center, the field

offers the “interesting ability to personalize medicine.”

Toxicogenomics tracks simultaneously the response of thousands of genes to environmental stimuli using glass slides of DNA combined with computational data analysis of the genes. The NIEHS has developed an innovative microarray technology called the ToxChip, and is building a library of ToxChip patterns that will eventually represent all known toxicants. Such tools should give environmental health researchers the ability to identify



## Environmental Genome Project: A Positive Sequence of Events

Why does one person develop disease following exposure to a certain environmental agent but another person doesn't? That question lies at the heart of the NIEHS's Environmental Genome Project (EGP), begun in late 1997. To find an answer, the EGP is studying variations of certain genes and their interactions with the environment. Scientists believe that if they can understand what comprises genetic susceptibility to environmental agents, they will be able to determine or predict what

the true risks of exposures are, and susceptible individuals can be better protected.

The scientific community has identified some 500 genes as being responsive to environmental factors such as diet and exposure to particular exogenous agents. The EGP is especially interested in the genes in two specific categories, DNA repair and cell cycle control, although these are not the only types of genes being studied. Currently, the EGP sponsors genome centers at the University of Washington and the University of Utah where much of the resequencing research is being conducted.

The first step in understanding the provenance of what have been called “environmentally related diseases” is to resequence environmentally responsive

genes to determine their possible allelic variations (single nucleotide polymorphisms, or SNPs). SNP variations play a key role in explaining why people react in different ways to different exposures.

Perhaps the most exciting product of the EGP to date is a central database of gene SNPs—the only one of its kind—which presents the information gathered through the EGP on an easy-to-use Web site. Housed at the Utah Genome Center, the database is accessible through the GeneSNPs link on the EGP home page at <http://www.niehs.nih.gov/envgenom/>. The database is constantly being updated and currently contains information on 485 different genes. Entries for each SNP include its length, genomic position, and frequency, as well as a model of the gene.

people who are at particular risk for being harmed by specific toxicants by enabling them to study how genes respond to particular chemicals. This may allow scientists to predict who might be susceptible to developing a particular disease and what the potential adverse responses might be.

The research produced by toxicogenomic studies promises to contribute to advancing intervention and prevention approaches to environmental diseases. The NCT will work closely with the

Environmental Genome Project to refine gene expression study techniques to help pinpoint variances in genetic susceptibility among people. NIEHS researchers are in the process of resequencing DNA repair enzyme and cell cycle control genes. Center projects will further work toward expanding the knowledge base on proteomics, or the micro-level study of proteins.

The NCT's database will eventually include the 20 years of toxicologic data contained in the National Toxicology

Program archives. Van Houten says that although not all the data have been entered into a computer yet, the ultimate goal is to link this information to the NCT's database.

Toxicogenomics will also help guide federal agencies and legislators to develop guidelines and laws that regulate the concentrations of various chemicals in the environment. Guidelines based on firm scientific data will provide a strong basis for health-related policy and regulation.

—Lindsey A. Greene



The next step in understanding the relationship between genes, environment, and disease is functional analysis—understanding the biologic significance of gene variations, and how variations cause their respective effects. Under the auspices of the EGP, grants for up to five Comparative Mouse Genomics Centers across the United States and Europe will be made by April 2001. The centers will develop mouse models that mimic human gene variations in order to examine the functional implications of these variations.

In addition to resequencing and functional analysis studies, the EGP funds population-based studies, which examine how different SNPs are distributed through various U.S. subpopulations and how they affect their carriers in terms of increased

disease susceptibility or resistance. The NIEHS Division of Extramural Research and Training is in the process of developing programs to distribute planning grants for molecular epidemiologists. Once SNPs are determined, the EGP will need people who can assess the frequency of SNP expression in the population. Molecular epidemiologists will be trained to effectively translate the data currently being generated by the EGP into population-based research.

Other components of the EGP include biostatistics/bioinformatics projects to develop statistical and computer models for analyzing gene–environment interactions, and to develop means for analyzing data on macromolecular cellular components such as DNA and proteins. In addition, technology development projects will

create better, faster ways to conduct EGP studies. Finally, ethical, legal, and social implications projects look at developing policies to address the complex issues raised by the ability to delve deeply into an individual's DNA.

The NIEHS is also involved in an interagency agreement with Lawrence Livermore National Laboratory to resequence a subset of 40–50 genes to determine the major background SNPs in the U.S. population. To date, resequencing has been completed for 15 genes among 90 disease-free people representing the racial distribution of the United States, with 18 more in progress. The project will next search the DNA of people with particular disease for SNPs that may be implicated in disease causation. —Susan M. Booker